Graph Regularized Matrix Factorization for MiRNA-Disease Association Prediction

Negin Fraidouni
Department of Computer Science and Engineering
The University of Texas at Arlington
Arlington, TX, 76010
neginfraidouni@gmail.com

Gergely Zaruba
Department of Computer Science and Engineering
The University of Texas at Arlington
Arlington, TX, 76010
zaruba@uta.edu

Abstract—MicroRNAs (miRNAs) are a small, non-coding class of RNAs; they are involved in the development and progression of many human diseases. Although many miRNA-disease associations have already been discovered, there are many more which are still unknown. Unfortunately, experimental verification of miRNA-disease associations is very expensive and time consuming. So, computational methods and bioinformatics algorithms can be applied to help scientists pinpoint the most likely associations for more experimental verification, thus making such future discoveries less time and energy consuming. In this paper we investigate the Graph Regularized Matrix Factorization (GRMF) method for miRNA-disease prediction. This method combines miRNA functional similarity, disease semantic similarity, and known miRNA-disease associations to determine the likelihood of unknown miRNA-disease associations. Using 6-fold cross validation, we show that the GRMF method can reach a mean AUC (area under the curve) of 0.91, outperforming three state-of-the-art methods. To test the performance of GRMF for diseases with no known associations, we used Breast Neoplasm, removing all related miRNAs; the 50 predicted miRNAs by GRMF were verified by the databases: HMDD v.2.0, dbDEMC, and miR2Disease. For another case study, we used Lymphoma using known associations from HMDD v.2.0; 45 out of 50 (90%) of the GRMF predicted miRNAs were verified by dbDEMC, miR2Disease and PubMed literature. Therefore, we believe that GRMF could be an effective method to predict miRNA-disease associations.

Keywords: miRNA-Disease Association; miRNA Functional Similarity; Disease Semantic Similarity; Matrix factorization

I. INTRODUCTION

MiRNAs are small (about 21 to 24 nucleotides), non-coding, single strand RNA molecules; they are involved in gene expression regulation. MiRNAs are found in most eukaryotes including those of humans, and they tend to bind target mRNA and prevent protein production [1]. MiRNA-directed gene expression regulation is a very active area of research. Hundreds of miRNAs have been discovered and the recent development of sequencing techniques and bioinformatics prediction methods significantly enhanced our information about miRNAs, including possible functions and regulatory targets [2]–[4]. MiRNAs have been found to be responsible for different processes including cell death, cell proliferation, neural patterning, immunity, fat metabolism, and hematopoietic differentiation [5]. Computational methods for finding genes regulated by miRNAs have suggested that all these examples only represent a few samples and thus they can not describe the whole miRNA system [6].

Dysregulation of miRNAs have been shown to be the main reason of abnormal cell behavior and hence some human diseases. More and more miRNAs have been confirmed to be responsible for the development of human diseases [7], [8]. For instance studies confirmed that the miR-200 family has a strong association with breast cancer [9]; also, leukemia is one of the human cancers confirmed to be related to and miR-16 dysregulation [10]. So recognizing miRNA-disease associations can help in diagnosing, treating, and preventing human diseases. However, it is prohibitive to find the associations one-by-one due to the significant amount of resources that have to be spent in performing such experiments. Meanwhile, known miRNA-disease associations are stored in databases like HMDD v.2.0 [11], dbDEMC [12] and miR2Disease [13] but there is a high demand for identifying new associations. Using computational methods to prioritize potential miRNAs for any specific miRNA-disease study could significantly reduce the time and financial resources needed for these experiments. Many computational methods that have been developed by scientists, are based on the assumption that similar miRNAs are likely to be related to similar diseases [14]–[16].

In this paper we will investigate how the Graph Regularized Matrix Factorization (GRMF) method could be used to discover miRNA-disease relationships. The rest of the paper is organized as follows: Section 2 describes the GRMF method in detail. In Section 3 we describe our GRMF experiments and compare our result with three state-of-the-art methods for predicting miRNA-disease associations. Finally, Section 4 concludes the paper.

II. METHODS

In this section, we first describe the datasets used, then we provide a description of Graph Regularized Matrix
Factorization as it applies to miRNA-disease associations.

A. Human miRNA-Disease Association
The database we have used for our study contains data of the associations between human miRNA and disease from the Human microRNA Disease Database (HMDD [11]). The database includes about 579 miRNAs, 384 diseases and 10,381 experimentally confirmed associations between miRNAs and diseases. Using this data, we can construct matrix $Y$ to capture the associations between miRNAs and diseases. Each row of matrix $Y$ represents a different miRNA and each column represents a different disease. Based on the datasets, the elements of matrix $Y$ can only be either 0 or 1. If miRNA $m_i$ is associated with disease $d_j$, then $Y_{ij}$ is 1. $Y_{ij}$ is 0 in the cases where there is no known association between $m_i$ and $d_j$ (this does not mean that there is no relationship but merely that the relationship is unknown).

B. miRNA Functional Similarity
miRNA functional similarity scores are calculated under the assumption that if two miRNAs are functionally similar, they are more likely to be related to phenotypically similar diseases. Wang et al. developed a method called MISIM [17] for measuring the similarity between two different miRNAs. MISIM has 4 main steps: first, diseases associated with two miRNAs are recognized, denoted as $d_1$ and $d_2$. In the next step the semantic values of diseases were calculated. In the third step the semantic similarity were calculated between every pair of the diseases between $d_1$ and $d_2$. And finally the functional similarity of two miRNA were calculated based on the semantic similarity of $d1$ and $d2$. We downloaded the scores from: http://www.cuilab.cn/files/images/cuilab/misim.zip

C. Disease Semantic Similarity
For the purpose of calculating disease semantic similarities, diseases can be described as a Directed Acyclic Graph (DAG). Each disease represents a node in the graph while edges represent relationship between diseases. Disease $d_1$ can be described as $DAG(d_1) = (d_1, S_{d_1}, E_{d_1})$ where $S_{d_1}$ is the set of all nodes including all ancestors of node $d_1$ including $d_1$ itself and $E_{d_1}$ is the set of all corresponding links, this includes all direct edges from parents to child nodes. So the contribution of disease $d$ in disease $d_1$ can be calculated as:

$$S_{d_1}(d) = \begin{cases} 1 & d_1 = d \\ \max \{ \Delta \ast S_{d_1}(d') \mid d' \in \text{children of } d \} & d_1 \neq d \end{cases}$$

(1)

Here $\Delta$ is the contribution factor for all connection links from disease $d$ to disease $d'$. The contribution of disease $d_1$ to its own semantic value is 1 so the farther the nodes from disease $d_1$, the less effect they have on the $d_1$’s semantic value. Thus, the value of $\Delta$ should be between 0 and 1. Wang et al. suggested that when the value of $\Delta = 0.5$ then the results show better correlation with the expression similarity [17]. The semantic value($SV$) of disease $d_1$ can be described as:

$$SV(d_1) = \sum_{d \in D_{d_1}} S_{d_1}(d)$$

(2)

If two diseases have much in common in the DAG then their similarity value would become larger. The semantic similarity between diseases $d_1$ and $d_2$ can be calculated as:

$$SV(d_1, d_2) = \frac{\sum_{d \in \{d_1, d_2\}} (S_{d_1}(d) + S_{d_2}(d))}{SV(d_1) + SV(d_2)}$$

(3)

where $S_{d_1}(d)$ is the semantic value of disease $d$ related to disease $d_1$ and $S_{d_2}(d)$ is the semantic value of disease $d$ related to disease $d_2$. What equation 3 calculates, is the semantic similarity between two different diseases based on their location in DAG and the common links in their ancestors.

D. Weighted K-nearest Known Neighbors
Our miRNA-disease association matrix $Y \in \mathbb{R}^{n \times m}$ has $n$ rows representing miRNAs and $m$ columns representing diseases. Matrix $Y$ is a sparse matrix and most of it’s values are zero although many of these zeros are unknown interactions that could potentially be true. Our aim is to replace zeros with a continuous value between 0 and 1; in the preprocessing step, we use the weighted k-nearest known neighbor algorithm to estimate an association likelihood based on the known associations. Algorithm 1 describes the process in detail. In a nutshell, first we calculate the weighted average of the k nearest neighbors to miRNA $m_i$, then we calculate the weighted average of the k nearest neighbors to disease $d_j$; in the final step we replace the entries in $Y$ that are 0 by the average likelihood of $m_i$ and $d_j$.

---

Algorithm 1: Weighted K-nearest Known Neighbors

**Input:** $Y$, $S^m$, $S^d$, $k$, $\eta$

$Y_m = Y_d = 0$

for $q = 1 \rightarrow n$:

\[\text{knn} = \text{k-nearest neighbors of row } q \text{ from } S^m\]

for $i > k$:

$w_i = \eta^{i-1} \ast S^m(q, \text{knn}_i)$

end for

$P^q_i = \sum_{i=1}^k S^m(q, \text{knn}_i)$

$Y_m(q) = \frac{1}{P^q_i} \sum_{i=1}^k w_i Y(q, \text{knn}_i)$

end for

for $r = 1 \rightarrow m$:

\[\text{dnn} = \text{k-nearest neighbors of row } r \text{ from } S^d\]

for $j > k$:

$w_j = \eta^{j-1} \ast S^d(r, \text{dnn}_j)$

end for

$P_r = \sum_{j=1}^k S^d(r, \text{dnn}_j)$

end for
E. Graph Regularized Matrix Factorization

A linear approximation of our miRNA-disease association matrix \( Y \in \mathbb{R}^{n \times m} \) can be shown by \( Y \approx W H^T \) where \( W \in \mathbb{R}^{n \times f}, H \in \mathbb{R}^{m \times f} \) and \( f \) is the number of latent features in \( W \) (in miRNAs) and \( H \) (in diseases).

Given a data matrix \( Y \), the choice of \( W \) and \( H \) have to be such to minimize the reconstruction error between \( Y \) and \( W H^T \). Among the various error functions that have been proposed \( [18] \), the most widely used is the squared error or euclidean distance with respect to the Frobenius norm. So the problem can be written as:

\[
\min_{W,H} \| Y - W H^T \|_F^2 \tag{4}
\]

The objective function in Eq. 4 is convex in \( W \) only or \( H \) only, but it is not convex in both of them together. For the aim of preventing overfitting, we can add linear and graph regularization terms. Linear regularization term minimizes norms of both \( W \) and \( H \) while graph regularization terms minimize the distance between latent feature vectors of two neighbor miRNAs and diseases. So our objective function becomes:

\[
\min_{W,H} \| Y - W H^T \|_F^2 + \lambda_a (\| W \|_F^2 + \| H \|_F^2) + \lambda_b \sum_{i=1}^n \sum_{q=1}^n (S^m_{i,q})^2 \| w_i - w_q \|^2 + \lambda_c \sum_{j=1}^m \sum_{r=1}^m (S^d_{j,r})^2 \| h_j - h_r \|^2
\]

where \( \lambda_a, \lambda_b \), and \( \lambda_c \) are all positive parameters, \( w_i \) and \( w_q \) are \( i^{th} \) and \( q^{th} \) row of \( W \), \( h_j \) and \( h_r \) are \( j^{th} \) and \( r^{th} \) row of \( H \). We can rewrite Eq. 5 as:

\[
\min_{W,H} \| Y - W H^T \|_F^2 + \lambda_a (\| W \|_F^2 + \| H \|_F^2) + \lambda_b \text{Tr}(W^T L_b W) + \lambda_c \text{Tr}(H^T L_c H) \tag{6}
\]

where \( \text{Tr} \) is the trace of a matrix, \( L_b = D^b - S^m \) and \( L_c = D^c - S^d \) are the graph Laplacians for \( S^m \) and \( S^d \) respectively and \( D^a_{i,i} = \sum_q S^m_{i,q} \) and \( D^a_{j,j} = \sum_r S^d_{j,r} \) are diagonal matrices. (We refer the reader to \( [19] \) for more details on obtaining Eq. 6 from Eq. 5.)

We provide a pseudocode for the Graph Regularized Matrix Factorization (GRMF) in Algorithm 2. We use singular value decomposition (SVD) to obtain \( U \in \mathbb{R}^{n \times f}, \Sigma \in \mathbb{R}^{f \times f} \) and \( V \in \mathbb{R}^{m \times f} \) from \( Y \). Then we initialize \( W \) and \( H \) as \( W = U \sqrt{\Sigma} \) and \( H = V \sqrt{\Sigma} \). We used alternating least squares to update \( W \) and \( H \) in each iteration. If we denote the objective function of Eq. 6 as \( J \), we set \( \frac{\partial J}{\partial W} = 0 \) and \( \frac{\partial J}{\partial H} = 0 \) so we can update \( W \) and \( H \) through:

\[
W = (YH - \lambda_b L_b W)(H^T H + \lambda_a I_k)^{-1}
\]

\[
H = (Y^T W - \lambda_c L_c H)(W^T W + \lambda_a I_k)^{-1}
\]

Algorithm 2: Graph Regularized Matrix Factorization (GRMF)

\[
\text{Input: } Y, S^m, S^d, f, \lambda_a, \lambda_b, \lambda_c
\]
\[
U \Sigma V^T = \text{SVD}(Y,f)
\]
\[
W = U \sqrt{\Sigma}
\]
\[
H = V \sqrt{\Sigma}
\]
\[
L_b = D^b - S^m
\]
\[
L_c = D^c - S^d
\]
\[
\text{while not converged:}
\]
\[
W = (YH - \lambda_b L_b W)(H^T H + \lambda_a I_k)^{-1}
\]
\[
H = (Y^T W - \lambda_c L_c H)(W^T W + \lambda_a I_k)^{-1}
\]
\[
\text{end while}
\]
\[
YY = WH
\]
\[
\text{Output: } YY
\]

III. EVALUATION

To evaluate the proposed GRMF-based method we compared its performance to three state-of-the-art miRNA-disease prediction methods.

A. Competitive Methods

1) RLSMDA: Chen et al. \([14]\) developed the method of Regularized Least Squares for MiRNA-Disease Association (RLSMDA) to find miRNAs associated with different diseases using a semi-supervised learning method. RLSMDA is designed using a continuous classification function to reflect the probability with which each miRNA is associated with a specific disease. RLSMDA can predict miRNAs related to diseases that have no known associated miRNA and it does not need negative miRNA-disease associations. However the ways of combining classifiers in different spaces and also the choice of parameters can affect the prediction performance of this method.

2) NetCBI: In this study Chen et al. \([15]\) constructed the miRNA-disease association network (NetCBI) using a representation of for bipartite graphs, where the nodes correspond to either diseases or miRNAs, and edges correspond to the associations between them. The main idea behind NetCBI is that if a given miRNA is related to a disease, other miRNAs that are similar to it, will be chosen and recommended to be related to that disease as well. Also if a miRNA is related to a disease, that miRNA will also be likely classified to be related to similar diseases.
3) **NBI**: Li et al. [16] developed a computational method (NBI) to predict new miRNA-disease associations by integrating environmental factor (EF) similarity and disease phenotypic similarity. More precisely, in NBI, three comprehensive bipartite networks are constructed, i.e., the EF-disease, the EF-miRNA, and the miRNA-disease associations. This method uses known associations to obtain predicted candidates. The miRNAs that are related to EFs, average their resources to all of their neighbors and thus they distribute the associations to every miRNA neighbor.

**B. Performance**

We plotted Receiver Operating Characteristics curve (ROC) and used Area Under the ROC curve (AUC) as the main metric for evaluating their performance. The area under ROC curve is calculated as an index of the prediction power of the GRMF method. The value of AUC is between 0 and 1 and higher amounts shows more prediction power. If the value is equal to 0.5, it means the performance is equal to a random prediction.

For an specific disease \(d_1\), all the known \(d_1\) related miRNAs are defined as labeled nodes and the remaining miRNAs (on which there is no relevance information) are defined as unlabeled nodes. Given a threshold \(\delta\), if the result prediction of a labeled node is greater than \(\delta\), then the node is identified as positive sample. If the result prediction of a an unlabeled node is less than \(\delta\), then the node is a considered as a negative sample.

To plot a ROC curve we calculated the true positive rates (TPR or sensitivity) and false positive rates (FPR, 1-specificity) through:

\[
TPR = \frac{TP}{TP + FN}, \quad FPR = \frac{FP}{TN + FP} \tag{9}
\]

where:

- \(TP\) : Number of correctly identified positive samples.
- \(TN\) : Number of correctly identified negative samples.
- \(FP\) : Number of misidentified positive samples.
- \(FN\) : Number of misidentified negative samples.

Sensitivity means the percentage of the positive samples that are correctly identified among all the positives and specificity means the percentage of the negative samples correctly identified among all negatives.

We conducted five repetitions of a 6-fold cross validation for each of the methods. The 6-fold cross validation is implemented using the known miRNA-disease association in the HMDD V2.0 database. So in each repetition, we divided our association matrix \(Y\) in to six parts and each parts, one-by-one , was left out as the test set while we used the remaining five parts as the training set. All of unknown miRNA-disease association pairs can be seen as candidate samples. After applying GRMF, scores of the test samples were compared with the all scores of the candidates samples. In order to make the validation more accurate, we repeated this process 5 times. Figure 1 depicts the ROC curve and the calculated AUC of each fold in 6-fold cross validation for GRMF method.

![Figure 1. Performance of each fold in 6-fold cross validation for GRMF method.](image)

We compare the performance of GRMF approach with three state-of-the-art methods for miRNA-disease association prediction. Figure 2 shows the prediction performance of GRMF, NBI [16], RLSMDA [14] and NetCBI [15]. The GRMF achieves the AUC value of 0.91 compared with other methods: NBI: 0.77, RLSMDA: 0.80 and NetCBI: 0.82 and outperforms the other three.

**C. Case studies**

In order to demonstrate the performance of GMRF, we evaluated the prediction ability of GMRF for miRNAs related to Breast Neoplasm and Lymphoma. Two miRNA-disease datasets (dbDEMC [12] and miR2Disease [13]) and previous PubMed studies were used to confirm the correctness of the prediction.

1) **Breast Neoplasm**: Breast Neoplasm (BN or breast cancer) is the second most common cancer in American women; dysregulation of miRNAs play an important role in this disease [22]. We used BN in order to show the performance of GRMF for diseases which have no related
miRNAs. The total number of miRNAs related to BN was 202 so we removed all 202 related miRNAs in our dataset to ensure that only the information from other diseases would be used to predict the related miRNAs to BN. We then ranked the predicted scores for all candidate miRNAs so the top 50 miRNAs was selected and they are shown in Table 1. Based on our results, we could confirm all 50 miRNAs by miR2Disease, dbDEMC and HDMM.

2) Lymphoma: Lymphoma is recognized as the fifth most common cancer type and is cancer of lymphatic system (Blood B and T cells). It includes Hodgkin Lymphoma (HL) and Non-Hodgkin Lymphoma (NHL) [20]. B-cells Lymphoma is the most common type of NHL in the United states and worldwide. Because Lymphoma can be derived form B-cells at different stages of cell cycle, miRNAs can be both target genes and specific markers [21]. For the second case study, we chose Lymphoma and the results are summarized in Table 2. We could confirm 45 miRNAs out of 50 as shown by dbDEMC, miR2Disease and experimental literature in PubMed.

IV. CONCLUSION

Finding the molecular mechanism of diseases can help exploring disease pathogenesis and finding effective treatments. MiRNAs as a class of non-coding RNAs are responsible for regulating gene expression so they can cause various diseases [23], [24]. Some computational approaches have been proposed to capture miRNA-disease association [14]–[16]. However, these methods have limitations.

In this paper we presented a Graph Regularized Matrix Factorization method (GRMF) to predict miRNA-disease associations based on the assumption that similar miRNAs (functionally) tend to be related to similar diseases (phenotypically). We used miRNA functional similarity, disease semantic similarity, and known miRNA-disease associations form the HDMM v.2.0 database. To verify the accuracy of the GRMF method, we used five repetitions of 6-fold cross validation. We compared the result of the GRMF method with three state-of-the-art methods and concluded that GRMF outperforms the other three in terms of AUC.

We selected Breast Neoplasm as a case study in order to show the performance of GRMF for diseases which have no related miRNAs and based on the results, we could confirm all 50 miRNAs as identified by miR2Disease, dbDEMC and HDMM. As the second case study we chose Lymphoma to demonstrate the performance of GRMF and based on the results, we could confirm 45 miRNAs out of 50 as identified by dbDEMC, miR2Disease and experimental literature in PubMed. The GRMF method could provide an effective approach to study miRNA-disease associations. We also recognize that GRMF has some limitations which can be improved in future research. For example, the sequence information of miRNAs is used to measure miRNA similarity but some studies show that the structural information...
Table I  
PREDICTION OF THE TOP 50 PREDICTED miRNAs ASSOCIATED WITH BREAST NEOPLASM BASED ON THE KNOWN ASSOCIATIONS IN HDMM V.2.0 DATABASE.

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Evidence</th>
<th>miRNA</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-mir-10b</td>
<td>dbDEMC; miR2Disease; HMDD</td>
<td>hsa-mir-210</td>
<td>dbDEMC; miR2Disease; HMDD</td>
</tr>
<tr>
<td>hsa-mir-372</td>
<td>dbDEMC</td>
<td>hsa-mir-516a</td>
<td>HMDD</td>
</tr>
<tr>
<td>hsa-mir-134</td>
<td>dbDEMC</td>
<td>hsa-mir-146b</td>
<td>dbDEMC; miR2Disease; HMDD</td>
</tr>
<tr>
<td>hsa-mir-143</td>
<td>dbDEMC; miR2Disease; HMDD</td>
<td>hsa-mir-192</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-let-7e</td>
<td>dbDEMC; HMDD</td>
<td>hsa-mir-183</td>
<td>dbDEMC; HMDD</td>
</tr>
<tr>
<td>hsa-mir-1</td>
<td>dbDEMC</td>
<td>hsa-mir-320a</td>
<td>HMDD</td>
</tr>
<tr>
<td>hsa-mir-499a</td>
<td>HMDD</td>
<td>hsa-mir-499a</td>
<td>HMDD</td>
</tr>
<tr>
<td>hsa-mir-150</td>
<td>dbDEMC</td>
<td>hsa-mir-182</td>
<td>dbDEMC; miR2Disease; HMDD</td>
</tr>
<tr>
<td>hsa-let-7i</td>
<td>dbDEMC; miR2Disease; HMDD</td>
<td>hsa-mir-152</td>
<td>dbDEMC; miR2Disease; HMDD</td>
</tr>
<tr>
<td>hsa-mir-137</td>
<td>dbDEMC; HMDD</td>
<td>hsa-mir-221</td>
<td>dbDEMC; miR2Disease; HMDD</td>
</tr>
<tr>
<td>hsa-mir-19b</td>
<td>dbDEMC</td>
<td>hsa-mir-184</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-125a</td>
<td>dbDEMC; miR2Disease; HMDD</td>
<td>hsa-mir-32</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-214</td>
<td>dbDEMC</td>
<td>hsa-mir-325</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-302b</td>
<td>dbDEMC</td>
<td>hsa-mir-30b</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-219</td>
<td>dbDEMC</td>
<td>hsa-mir-106a</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-204</td>
<td>dbDEMC; miR2Disease; HMDD</td>
<td>hsa-mir-205</td>
<td>dbDEMC; miR2Disease; HMDD</td>
</tr>
<tr>
<td>hsa-mir-20b</td>
<td>HMDD</td>
<td>hsa-mir-181c</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-101</td>
<td>dbDEMC; miR2Disease; HMDD</td>
<td>hsa-let-7b</td>
<td>dbDEMC; HMDD</td>
</tr>
<tr>
<td>hsa-mir-302c</td>
<td>dbDEMC</td>
<td>hsa-mir-212</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-20a</td>
<td>miR2Disease; HMDD</td>
<td>hsa-mir-506</td>
<td>HMDD</td>
</tr>
<tr>
<td>hsa-let-7d</td>
<td>dbDEMC; miR2Disease; HMDD</td>
<td>hsa-mir-140</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-25</td>
<td>dbDEMC</td>
<td>hsa-mir-708</td>
<td>HMDD</td>
</tr>
<tr>
<td>hsa-mir-195</td>
<td>dbDEMC; miR2Disease; HMDD</td>
<td>hsa-mir-433</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-107</td>
<td>dbDEMC</td>
<td>hsa-mir-153</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-187</td>
<td>dbDEMC</td>
<td>hsa-mir-141</td>
<td>dbDEMC</td>
</tr>
</tbody>
</table>

Table II  
PREDICTION OF THE TOP 50 PREDICTED miRNAs ASSOCIATED WITH LYMPHOMA.

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Evidence</th>
<th>miRNA</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-mir-141</td>
<td>dbDEMC</td>
<td>hsa-mir-182</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-10b</td>
<td>dbDEMC</td>
<td>hsa-let-7e</td>
<td>dbDEMC; miR2Disease</td>
</tr>
<tr>
<td>hsa-mir-30a</td>
<td>dbDEMC</td>
<td>hsa-mir-335</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-193b</td>
<td>PMID:22235305</td>
<td>hsa-mir-183</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-151a</td>
<td>Unconfirmed</td>
<td>hsa-mir-148a</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-106a</td>
<td>dbDEMC; miR2Disease</td>
<td>hsa-mir-34a</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-221</td>
<td>dbDEMC</td>
<td>hsa-mir-9</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-7</td>
<td>dbDEMC</td>
<td>hsa-mir-125b</td>
<td>PMID:23527180</td>
</tr>
<tr>
<td>hsa-mir-195</td>
<td>dbDEMC</td>
<td>hsa-mir-429</td>
<td>Unconfirmed</td>
</tr>
<tr>
<td>hsa-mir-214</td>
<td>dbDEMC</td>
<td>hsa-mir-100</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-29b</td>
<td>dbDEMC</td>
<td>hsa-mir-132</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-219b</td>
<td>dbDEMC</td>
<td>hsa-let-7i</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-196a</td>
<td>dbDEMC</td>
<td>hsa-mir-205</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-378a</td>
<td>Unconfirmed</td>
<td>hsa-mir-192</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-191</td>
<td>dbDEMC</td>
<td>hsa-mir-223</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-133a</td>
<td>dbDEMC</td>
<td>hsa-mir-30e</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-103a</td>
<td>Unconfirmed</td>
<td>hsa-mir-145</td>
<td>dbDEMC; miR2Disease</td>
</tr>
<tr>
<td>hsa-mir-146b</td>
<td>PMID:24931464</td>
<td>hsa-mir-194</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-30C</td>
<td>dbDEMC</td>
<td>hsa-mir-15b</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-34b</td>
<td>dbDEMC</td>
<td>hsa-mir-142</td>
<td>Unconfirmed</td>
</tr>
<tr>
<td>hsa-mir-152</td>
<td>dbDEMC</td>
<td>hsa-mir-30d</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-26b</td>
<td>dbDEMC</td>
<td>hsa-mir-143</td>
<td>dbDEMC; miR2Disease</td>
</tr>
<tr>
<td>hsa-mir-338</td>
<td>dbDEMC</td>
<td>hsa-mir-22</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-29a</td>
<td>dbDEMC</td>
<td>hsa-mir-96</td>
<td>dbDEMC</td>
</tr>
<tr>
<td>hsa-mir-27b</td>
<td>dbDEMC</td>
<td>hsa-mir-222</td>
<td>dbDEMC</td>
</tr>
</tbody>
</table>

can be more effective [25], [26]. Furthermore, expression information of miRNAs could also be used to measure this similarity.

REFERENCES


